

## Article

# Detecting Risk Factors of Road Work Zone Crashes from the Information Provided in Police Crash Reports: The Case Study of Portugal

Bertha Santos <sup>1,2,\*</sup> , Valdemiro Trindade <sup>1</sup>, Cláudia Polónia <sup>1</sup> and Luís Picado-Santos <sup>2</sup> 

<sup>1</sup> Department of Civil Engineering and Architecture, University of Beira Interior, Calçada Fonte do Lameiro, 6200-358 Covilhã, Portugal; vmarquestrindade@gmail.com (V.T.); claudiascpolonia@gmail.com (C.P.)

<sup>2</sup> CERIS, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisbon, Portugal; luispicadosantos@tecnico.ulisboa.pt

\* Correspondence: bsantos@ubi.pt

**Abstract:** Several studies have shown that European police crash reports provide different detail degrees of work zone crash-related data. In this sense, the present study aims to verify the possibility of identifying significant risk factors involved in the occurrence of road work zone crashes with casualties, based on the official data usually available, through a descriptive, binary logistic, and probit regression statistical analysis. To accomplish the analysis, a total of 2597 police-reports related to 1767 Portuguese work zone crashes that occurred during the 2013–2015 period were considered and binary logistic and probit regression models were estimated by the main type of crash, contributing factor, and driver age group. Fifteen explanatory variables, selected based on the literature review and crash data provided in police crash reports, were considered in the analysis. The results obtained for the estimated coefficients and goodness-of-fit test values were found very similar for both link functions (logit and probit) and it was possible to identify risk factors. The modeling results pointed to excessive speed, disregard for vertical signs, luminosity, intersections, and motorcycle and heavy vehicle involvement as the most significant risk factors. Given the results, it is possible to conclude that binary logistic regression can be used in the statistical analysis of the available police official work zone crash data to identify and get some insight into the risk factors involved in work zone crashes. Data analysis also revealed the need to promote adequate and complete crash report filling by police officers. While police crash reports are not revised and standardized to incorporate more detailed work zone crash information, this approach can be used to support a more efficient road operation decision making and the review of some aspects related to work zone layout design.

**Keywords:** road safety; work zone; police crash reports; crash data; logit and binary logistic regression



**Citation:** Santos, B.; Trindade, V.; Polónia, C.; Picado-Santos, L. Detecting Risk Factors of Road Work Zone Crashes from the Information Provided in Police Crash Reports: The Case Study of Portugal. *Safety* **2021**, *7*, 12. <https://doi.org/10.3390/safety7010012>

Academic Editor: Garrett Mattos

Received: 11 November 2020

Accepted: 30 January 2021

Published: 4 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The road infrastructure is an essential component of a country's development. To maintain its quality and safety standards, maintenance and reconstruction actions are needed. The performance of such actions makes it necessary to study the impact that roadworks may have on crashes as work zones change locally and temporarily the traffic conditions offered to drivers (lower speeds, presence of work equipment and workers, narrow lanes, changes in vertical and horizontal signs, etc.). These changes in traffic conditions may cause the occurrence of serious crashes and casualties, requiring additional attention from drivers to adapt themselves to the new driving conditions.

Several studies like the ones performed by Ullman et al. (2006), Meng et al. (2010), Silverstein et al. (2016), La Torre et al. (2017), and Ambros et al. (2020) concluded that the probability of serious accident occurrences at road work zones is higher than those outside these areas [1–3] and that an increase in the expected crash frequency during the time

a work zone is installed is verified when compared to the crash frequency in the before period [4,5].

A review of the road safety literature revealed that vehicle-vehicle angle, rear-end and run-off-road crashes, and vehicle running over a pedestrian (including road workers on foot), are the most common types of crashes in work zones [6–10]. Most of these crashes involved vehicles running straight and some of the most significant contributory factors are excessive speed for the road conditions, the presence of an unexpected obstacle on the road, and the disregard for traffic control devices and signs [6,10–13]. Most of the involved drivers belong to the middle age group (25 to 64 years) and the crash consequences are essentially slight casualties, being the run over pedestrians in zebra crossings and workers on the road the main action related to the pedestrian victims. In most cases, a more significant monthly period can be identified in the warmer months of the year, chosen for carrying out roadworks (climatic conditions allow for more effective and durable results); the presence of a police officer is pointed as a favorable factor (decreases the crash occurrence probability) [7,11], and the involvement of heavy vehicles is common [6,8,11,13].

Although in most countries the information available on work zone crashes is still scarce, it is generally possible to determine its representativeness in the total number of crashes. For instance, the European data used in the project FORMAT (FORMAT—Fully Optimized Road Maintenance), dated 2005 and which counted with the participation of 20 organizations of 14 European countries, including Portugal and the United States of America, revealed that work zone crashes represent about 2% of all road crashes [14]. When considering only the highway networks this value may reach 5% of total crashes. Similar values (2 to 3%) were also found in an earlier European project, ARROWS (ARROWS—Advanced Research on Road Work Zone Safety Standards in Europe) [15]. Both projects conclude that the official data stored by the participating countries present varying degrees of inconsistency and most do not include important information needed for adequate analysis and modeling support.

In line with these project findings, an analysis of several police crash reports with work zone related fields was carried out (from Canada, USA, Portugal, Spain, Italy, and the United Kingdom) [16–23] and revealed the existence of different degrees of detail in the work zone related information collected for statistical purposes. All the analyzed reports record data on the existence of road work zones, obstacles on the pavement, and casualties related to pedestrian workers on the road. The use of temporary road signs is not clearly reported in Portugal and the U.K., and the use of reflective material by workers on the road can be registered in Portugal, Spain, and Massachusetts (USA).

Table 1 presents the specific information that should be considered in the structure of a work zone safety database to allow an in-depth analysis of this type of crash [14,15,24–28]. As identified in several European projects and by the authors, only part of that information is collected in most of the police crash reports in use. Information related to work zone type, duration, size, layout, operation, and crash location in the work zone is still scarce or nonexistent.

Even with the limitations mentioned, a statistical analysis approach of the available official data can be considered to verify the possibility of detecting risk factors involved in work zones crashes.

As the occurrence of crashes can be expressed as a binary variable (occurrence/non-occurrence) and they are affected by various factors, several authors have used binary logistic and probit regressions as a suitable statistical method to address the related variables and explain the impact of factors on the occurrence of crashes [27,29–31]. This approach can be applied to the current work zone collected data to assess the relationship between the probability of the occurrence of crashes and explanatory variables.

In this sense, this study aims to verify the possibility of identifying, based on the official crash data currently available in most statistical bulletins, risk factors with impact in work zone crashes and to compare the results provided by two statistical methods. To achieve this goal, binary logistic and probit regressions were considered suitable statistical

methods to adopt to develop prediction models by main crash type, contributing factor, and driver age group.

**Table 1.** Information for analysis of crashes in work zones.

Topic	Field	Crash Report
Type	Construction, reconstruction, maintenance	Kentucky (USA)
Duration	Long-term, short-term, mobile	***
Size	Length of work zone	***
Work zone layout	Lane narrowing/closure, diversion/detour, contraflow/crossover, alternative one-way traffic	United Kingdom (*)
Work zone location	Intersection, shoulder/roadside, median, footway/bikeway, tramway	All
Speed	Temporary speed limit	Portugal, Spain, British Columbia (Canada)
Traffic control devices employed	Traffic signs, traffic markings, traffic lights, variable message signs (VMSs), other	Portugal, Spain, Kentucky (USA), Texas (USA), Massachusetts (USA), Italy (**)
Other road equipment	Closure equipment, warning equipment, guiding equipment, protective equipment, road reflectors, speed reducer bumps, other	Massachusetts (USA)
Enforcement/Publicity	Police presence, electronic devices/cameras, traffic information, other	Spain, Kentucky (USA), Texas (USA), British Columbia (Canada)
Work zone operation	In operation, not in operation, in process of being assembled/disassembled, abandoned	***
Crash location in work zone	Advance warning area, transition area, buffer zone, activity area, termination area	***

Notes: \* mention to temporary road layout; \*\* in most cases only work zone presence signs; \*\*\* Several crash reports, such as those used in Spain, the United Kingdom, Kentucky, Texas, Massachusetts, or British Columbia, include a specific field for a sketch, accident description or contributory factors (with the option "work zone" or similar). This field can be filled with information about the work area and crash characteristics, but there are no specific instructions about the type of information to be registered.

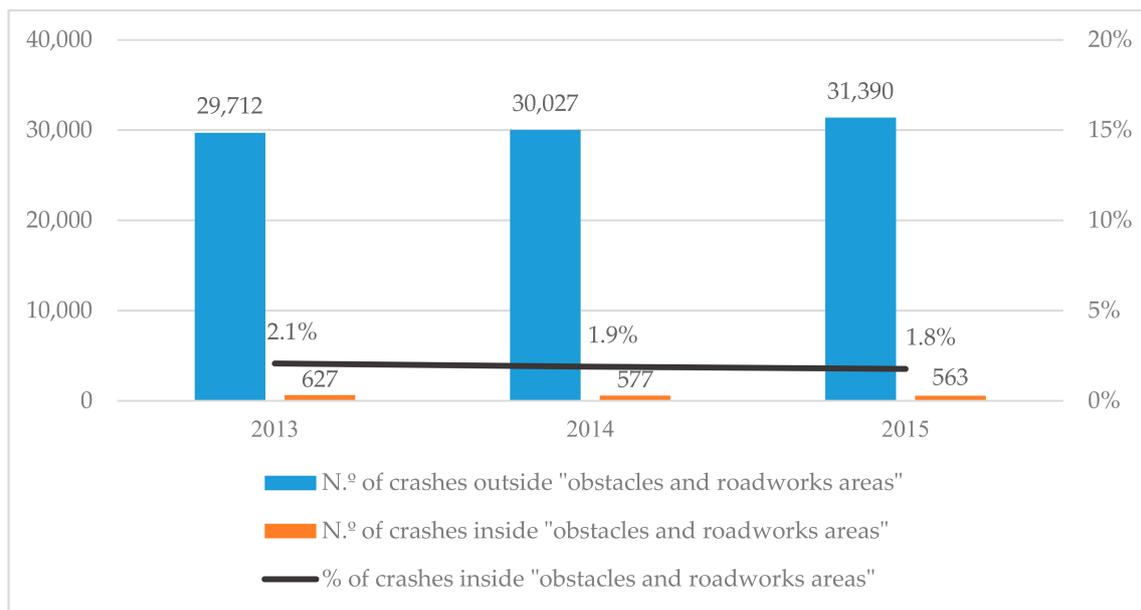
## 2. Method

The present study relies on the analysis of the official data related to work zone crashes involving casualties (fatal, serious, and slight) that occurred in mainland Portugal between 2013 and 2015 (not including the Azores and Madeira islands). The data, collected from the police crash reports, was provided by the Portuguese National Authority of Road Safety (ANSR) [32] and organized within the scope of two master's dissertations completed at the Department of Civil Engineering and Architecture of the University of Beira Interior (Portugal) [33,34]. The database that resulted from these studies was used in the scope of this research to assess the possibility of identifying risk factors associated with road work zone crashes from the information recorded in police crash reports.

Figure 1 shows the evolution of the number of work zone crashes in Portugal for the years considered in the analysis period. It should be noted that the decrease in work zone crashes does not necessarily mean an improvement in work zone safety. This is more likely to be related to the lower investment in Portuguese road construction and maintenance sectors verified in the period considered.

The data, organized for the diagnosis of the main aspects related to work zone crashes, allowed the identification of the most frequent types of work zone crashes and main contributing factors (descriptive analysis). Three driver age groups were also defined and the set of explanatory variables to be used in the statistical analysis was established.

With this information binary logistic and probit regression models were developed to identify the most relevant risk factors associated with the occurrence of work zone crashes and to predict its probability. The modeling analysis was performed using SAS/STAT® software (13.2, SAS®Institute Inc., Cary, NC, USA) [35] and IBM SPSS® Statistics software (24, IBM Corp., Armonk, NY, USA) [36].



**Figure 1.** Evolution of the number of crashes in Portuguese road work zones.

### 2.1. Crash Data, Descriptive Analysis, and Variables Selection

In Portugal, work zone crashes are recorded together with the crashes that occurred due to the presence of obstacles on the road pavement. The information is recorded in a specific field of the police crash report named "obstacles or roadworks areas" [20].

The data organization process revealed that the indication "Not set" was registered in this field in a significant number of reports, corresponding to a non-filled situation (22.1% of the crashes over the analysis period). Without this information, it is not possible to conclude whether the crash occurred or not in a work zone or due to the presence of an obstacle on the road pavement. To improve the reliability of data and results, these records were deleted from the database, and only correctly filled reports were considered in the analysis, corresponding to 1767 crashes.

Subsequently, two work zone crash datasets were structured. In the first, the information was organized by the crash and used for modeling the type of crash and main contributing factors (1767 crashes). Since the Portuguese official data does not identify the responsible driver for the crash occurrence, the second one was organized by the driver involved in the work zone crash and used for the age group modeling (2597 records/drivers). This aspect should be taken into account when interpreting results.

Following the data selection, a diagnosis of the main aspects related to work zone crashes were conducted to identify the most frequent type of crashes, the main contributing factors, and the distribution of drivers involved by age group (dependent variables). The main results obtained in the descriptive statistical analysis that supported the selection of the dependent variables are shown in Table 2.

Of the 17 types of crashes recorded in the Portuguese police crash report, those presented in Table 2 were the most frequent. However, with percentages between 8 and 10% of all work zone crashes, it is still possible to highlight the collision with vehicle or obstacle, the run off road without containment device, and the run off road with collision (with immobilized vehicle or obstacle). Regarding the contributing factors, the low percentages verified are related to a high number of non-filling situations confirmed for this field report (for 71.7% of all drivers), which limits the quality and the interpretation of model results. Despite this fact, the obtained results are very similar to the ones found in the reviewed literature [6–8,11,24,29].

**Table 2.** Descriptive statistics of the dependent variables.

Crash Type	% of WZ Crashes	Contributing Factors	% of WZ Crashes	Driver Age Group	% of Drivers Involved in WZ Crashes
Run off road (simple)	13.1%	Excessive speed	6.3%	Under 25 years old	13.8%
Pedestrian	12.6%	Unexpected obstacle (for at least one driver involved)	8.3%	Between 25 and 64	75.1%
Angle	11.9%	Disregard for vertical signs (for at least one driver involved)	2.9%	Over 64 years old	11.1%
Rear-end	10.8%	Disregard for safety distance (for at least one driver involved)	2.8%		
Others	51.6%	Others	79.7%		

Note: WZ means work zone.

A descriptive statistical analysis of the remaining data was also performed to support the selection of explanatory variables. Based on previous studies on the identification of work zone risk factors, available data, and descriptive statistical results, the fifteen explanatory variables presented in Table 3 were chosen and used in the risk factors identification analysis.

**Table 3.** Descriptive statistics of the explanatory variables used in models and assigned binary code.

Variable	Description	% Yes (1)	% No (0)
Urban	The crash occurred in an urban environment (not rural)	73.5%	26.5%
HV involved	At least one heavy vehicle involved	7.2%	92.8%
Moto involved	At least one motorcycle involved	25.5%	74.5%
Driver action	Before crash, at least one driver involved was running straight	75.1%	24.9%
Excessive speed	Before crash, and for the prevailing conditions, at least one driver involved exceeded speed	6.3%	93.7%
Obstacle	Unexpected obstacle on the road (for at least one driver involved)	8.3%	91.7%
Vertical signs	At least one driver involved disregarded vertical signs	2.9%	97.1%
Safety distance	At least one driver involved disregarded safety distance to the vehicle ahead	2.8%	97.2%
Speed limit	Posted speed limit $\geq$ 90 km/h (high standard divided roads—freeways)	7.2%	92.8%
Intersection	The accident occurred in a road intersection	25.8%	74.2%
Horizontal design	The road horizontal geometric design is straight (not curved)	73.8%	26.2%
Vertical design	The road vertical geometric design is level (not upgrade/downgrade)	64.2%	35.8%
Pav. Grip conditions	Clean and dry pavement (not with ice, snow, mud, gravel, sand, oil, or wet).	66.7%	33.3%
Luminosity	Good luminosity conditions—daylight (not dazzling sun, dawn, dusk, or night)	68.7%	31.3%
Weather	Clear (not rain, strong wind, fog, snow, smoke, or hail)	79.2%	20.8%

The literature review also revealed that several binary logit/probit regression crash studies consider dichotomous explanatory variables [29,30]. In this study, all explanatory variables were dichotomized. The assignment of 1 (occurrence) or 0 (non-occurrence) took into account the predominant factors (highest percentage) revealed by the descriptive statistical analysis of data (for variables urban, driver action, horizontal design, vertical design, pavement grip conditions, luminosity, and weather) and relevant work zone factors identified in precedent studies (for variables heavy vehicle (HV) involved, moto involved, excessive speed, obstacle, vertical signs, safety distance, speed limit, and intersection) [6–8,10–13].

Results of the descriptive statistical analysis and dichotomous variable code are presented in Table 3 (see columns % Yes (1) and % No (0)). The meaning of the variables coded with 1 is presented in the description field of the table.

## 2.2. Logistic and Probit Regression

Binary logistic regression is used to predict a dichotomous variable from a set of explanatory variables that can be nominal, continuous, or a mix of both. The primary reason why it is used is that the curve to describe the relationship between a set of predictors and the dependent variable, it not likely to be linear but rather an S-shape [37,38].

Considering that  $Y$  denotes an event ( $Y = 1$  and  $Y = 0$  denote the occurrence and nonoccurrence, respectively) and  $X$  a set of predictors  $\{X_1, X_2, \dots, X_k\}$ , then the occurrence probability of  $Y$  given  $X$  can be expressed by expression (1), where  $\{\beta_0, \beta_1, \dots, \beta_k\}$  are the estimated coefficient of predictors.

$$P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (1)$$

Since the dependent variable can only adopt two values (1 and 0), the logit is a link function used to linearize the dependent variable turning it into a variable with a linear behavior. Expression (1) can be transformed in a logit form as presented in expression (2).

$$\text{logit}(Y = 1|X) = \ln(\text{odds}) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (2)$$

In (2) the odds mean the ratio between the probability of a random event ( $P$ ) and its complementary event ( $1 - P$ ) or, in this case, the ratio between the probability of occurrence of a work zone crash due to a set of risk factors and the probability of nonoccurrence of the same event. A negative coefficient  $\beta$  indicates a negative influence in the event which reduces its probability of occurrence. On the other hand, a positive coefficient increases the probability of the event occurrence. In terms of values, a higher and lower value indicates respectively more or less influence in the occurrence of the event.

Another link function commonly used in models with dichotomous qualitative dependent variables is the normal distribution function, i.e., the probit model [38]. The link function and the adjusted model are presented in expressions (3) and (4), respectively.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha + \beta X} \exp\left(-\frac{1}{2}z^2\right) dz \quad (3)$$

$$\Phi^{-1}(\eta) = \beta_0 + \beta_1 X + \varepsilon \quad (4)$$

In these expressions  $\Phi$  is the normal distribution function;  $\alpha$  is a constant, break-event point, or threshold which defines the frontier region between the two values of  $Y$ ; and  $\beta X$  is the regression vector ( $\beta X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ ). The value of  $\beta X$  is taken to be the  $z$ -value of a normal distribution, where a higher value results in a larger area defined by the integral, meaning that the event is more likely to occur [37]. The probit model assumes that  $Y$  is the operationalization of a continuous latent variable  $\eta$  where  $\varepsilon$  represents the errors.

Comparing both methods, the logit transformation offers a great advantage in terms of understanding results, since it is easier to interpret the odds ratio than the estimated coefficients. For this reason, this transformation is largely adopted by researchers in different scientific fields. To compare the coefficients from both models the logit coefficients must be standardized dividing them by the logit standard deviation ( $\pi/\sqrt{3}$ ). The standardized logit coefficients are then directly comparable to the probit coefficients and the results of the two approaches can be compared.

Regarding the sample size, Long (1997) [39] suggests that maximum likelihood estimation including logistic regression must consider at least 100 cases, considers that 500 cases are generally adequate and that there should be at least 10 cases per predictor. Peduzzi et al. (1996) [40] refined this 10:1 recommendation, stating that ten times the number of predictors  $k$  should take into account the proportion  $p$  of successes ( $n = 10 k/p$ ). If the resulting number is less than 100 it should be increased to 100 as suggested by Long (1997).

In a more recent comparative study of logit and probit link functions for binary regression, Freitas et al. (2013) [41] recommend the use of the logit function for small-sized samples (<20) and conclude that for larger samples both functions present similar results.

### 3. Results

#### 3.1. Modeling Process

During the modeling process, the Maximum Likelihood Estimate was used to estimate the regression parameters  $\{\beta_0, \beta_1, \dots, \beta_k\}$  that maximizes the likelihood of the observed outcomes. Once the parameters were estimated, the probability of the event occurrence can be determined based on Equation (1).

The Wald, Likelihood-Ratio, and Score tests were used to test the significance of parameters of the overall model. It tests the null hypothesis that all coefficients of predictors are equal to zero ( $\beta_0 = 0, \beta_1 = 0, \dots, \beta_k = 0$ ). If significant at a 0.05 level, the null hypothesis is rejected, meaning that the predictors are influent in the prediction result. The Wald test was also applied to examine the significance of individual model parameters (explanatory variables).

Stepwise selection with a fast-backward elimination method was used in the regression process to remove statistically non-significant variables. The significance level adopted was 0.05 which means a more restrictive model with a smaller number of influent predictors. A model with too many predictors becomes a saturated model that does not describe the outcome variable properly.

Once the estimated values of coefficients are obtained, it is necessary to examine the goodness-of-fit of the model by measuring how well the model fits the observations. Hosmer–Lemeshow test was adopted to measure the goodness-of-fit. It tests the null hypothesis that there is no difference between the observed and predicted values of the response variable. If not significant at a 0.05 level, the null hypothesis cannot be rejected indicating that the model fits the data well [42].

#### 3.2. Results

Tables 4 and 5 show some examples of the modeling results obtained in this study, namely for angle crashes and the elderly driver age group.

**Table 4.** Estimated coefficients and odds for angle crash model.

Number of Observations		1767			
Response Profile		1	211		
		0	1556		
Link Function		Odds Ratio	Coefficient Logit	Coefficient Logit (standardized)	Coefficient Probit
Parameters	Intercept	-	-4.2082	-2.3201	-2.2831
	Urban	1.639	0.4940	0.2724	0.2628
	Driver action	2.314	0.8390	0.4626	0.4140
	Obstacle	0.296	-1.2166	-0.6707	-0.6120
	Vertical signs	3.922	1.3667	0.7535	0.8129
	HV involved	2.286	0.8267	0.4558	0.4695
	Intersection	4.828	1.5744	0.8680	0.8450
	Grip conditions	1.649	0.5000	0.2757	0.2413
Hosmer and Lemeshow Test			0.5065	0.4149	

**Table 5.** Estimated coefficients and odds for elderly driver involvement.

Number of Observations		2668			
Response Profile		1	288		
		0	2309		
Link Function		Odds Ratio	Logit	Logit (standardized)	Probit
Parameters	Intercept	-	-2.9791	-1.6425	-1.6814
	HV involved	0.141	-1.9613	-1.0813	-0.8992
	Driver action	1.385	0.3260	0.1797	0.1729
	Vertical signs	2.029	0.7077	0.3902	0.3886
	Grip conditions	1.470	0.3850	0.2123	0.1942
	Luminosity	1.641	0.4954	0.2731	0.2549
Hosmer and Lemeshow Test			0.9913	0.9905	

In a total of 1767 work zone crashes, 211 occurred by angle crash. Odds ratio (OR) results show that angle crashes in work zones are almost 5 times (4.828) more likely to occur inside a road intersection than outside. A clean and dry pavement, the involvement of heavy vehicles, the disregard for vertical signs, the driving straight action, and the urban environment are also contributing factors for these crashes ( $OR > 1$ ). When an unexpected obstacle is present on the road, the occurrence probability of angle crashes decreases ( $OR < 1$ ).

Comparing the estimated coefficients, it is possible to verify that the values are very close for both link functions, meaning that function choice does not significantly change the model description nor the predictor influence of the dependent variable occurrence (angle crashes). The Hosmer–Lemeshow goodness-of-fit test demonstrates a slightly higher value using the logit procedure, which means that the logit function fits the data more accurately.

Table 5 presents the results of the model estimation for crashes involving elderly aged drivers. In a total of 2597 drivers involved in 1767 work zone crashes, 288 are elderly aged (65 years old or older). The model shows that these drivers are more likely to be involved in work zone crashes when they drive straight under good grip and luminosity conditions and especially when the disregard for vertical signs is verified. The involvement of heavy vehicles had a negative impact in crashes involving elderly aged drivers ( $OR < 1$ ), meaning that work zone crashes involving older drivers were more likely to occur with the absence of heavy vehicles. The 11 remaining explanatory variables were not influential in these crashes. With a higher sample size (288 cases), estimated coefficients by probit and link functions were once again very close. The goodness-of-fit test also provided very similar results for both methods indicating a higher fit of the model to the data.

For an overall assessment, a summary of the qualitative impact of predictors in the probability of crash occurrence for the 11 developed models are presented in Tables 6–8, where columns indicate the different models and rows denote explanatory variables (predictors).

Table 6. Summary of model qualitative findings for crash type.

Crash Type Model	Pedestrian			Angle			Run of Road (Simple)			Rear End		
Sample Size	223	210	13	211	175	36	232	143	89	190	95	95
Factor	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural
Urban	↑*	n.a.	n.a.	↑	n.a.	n.a.	↓	n.a.	n.a.	↓	n.a.	n.a.
HV involved	-	-	-	↑	-	↑	↓*	-	↓	-	-	-
Moto involved	↓*	↓*	-	-	-	-	↑*	↑*	↑*	↓	↓	↓*
Driver action	↓	↓	↓	↑	↑	-	-	↑	-	↑	↑	-
Excessive speed	↓	↓	-	-	↓	-	↓	↓	-	↑**	↑**	-
Obstacle	↓	↓	-	↓*	↓*	-	-	-	-	↓*	↓*	↓
Vertical signs	-	-	-	↑**	↑**	↑*	-	↓*	-	-	-	-
Safety distance	↓	↓	-	-	-	-	-	-	-	↑*	↑*	↑*
Speed limit	-	-	-	-	-	-	-	-	-	-	-	-
Intersection	↓	↓	-	↑*	↑*	↑**	↓	↓	-	-	-	-
Horizontal design	↑**	↑*	-	-	-	-	↓	-	-	↑	↑	↑**
Vertical design	-	-	-	-	-	-	-	-	-	-	-	-
Pav. Grip conditions	↑	↑	-	↑	↑	-	↓	↓	-	-	-	-
Luminosity	↑	↑**	-	-	-	-	-	-	-	↑	-	↑
Weather	-	-	-	-	-	-	↑**	↑**	-	-	-	-
Hosmer and Lemeshow Test	0.455	0.989	1.000	0.506	0.216	0.145	0.077	0.619	1.000	0.634	0.648	0.949
Stepwise Number of steps	6	5	13	9	8	12	6	6	12	7	7	10
Cases correctly classified (%)	87.8	84.4	97.2	88.9	87.7	92.8	86.7	89.0	81.0	89.4	92.1	80.8
Classification by chance (%) <sup>(a)</sup>	77.9	72.9	94.6	79.0	76.7	85.8	77.2	80.4	69.2	80.8	86.4	67.7

Notes: This table summarizes the qualitative findings of the study, where “-” means that the factor is not influent in the model; “↑” means that the factor impact is likely to increase the probability of occurrence of certain work zone crash type, ↑\* and ↑\*\* are the two largest impact factors that likely increase the probability; “↓” means that the factor impact is likely to decrease the probability of occurrence; ↓\* is the smaller impact factors that likely decrease the probability and “n.a.” means that the factor was not considered in the model. <sup>(a)</sup> Proportional chance criterion.

**Table 7.** Summary of model qualitative findings for contributing factors.

Contributing Factors	Unexpected Obstacle			Excessive Speed			Disregard for Vertical Signs			Disregard for Safety Distance		
Sample Size	209	170	39	163	100	63	74	67	7	67	51	16
Factor	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural
Urban	↑**	n.a.	n.a.	-	n.a.	n.a.	↑	n.a.	n.a.	-	n.a.	n.a.
HV involved	-	-	-	-	-	-	↑**	↑**	-	-	-	-
Moto involved	↑*	↑*	-	-	-	-	-	-	-	↓*	↓*	-
Driver action	-	-	-	↓*	↓*	↓*	-	-	-	-	-	-
Excessive speed	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Obstacle	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Vertical signs	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Safety distance	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Speed limit	-	-	-	↑*	-	↑*	-	-	-	-	-	-
Intersection	↓	↓	-	-	-	-	↑*	↑*	-	-	-	-
Horizontal design	-	-	-	↓	↓	↓	-	-	-	↑*	↑**	↑*
Vertical design	↓	↓	-	-	-	-	-	-	-	-	-	-
Pav. Grip conditions	↓*	↓*	-	-	-	-	↑	↑	-	↑**	↑*	-
Luminosity	↓	↓	↓*	↓	-	-	-	-	-	-	-	-
Weather	-	-	-	-	-	-	-	-	-	-	-	-
Hosmer and Lemeshow Test	0.182	0.647	-	0.281	0.387	0.311	0.060	0.547	-	0.488	0.983	1.000
Stepwise Number of steps	6	5	10	8	8	8	8	7	11	9	7	9
Cases correctly classified (%)	88.2	86.9	91.7	90.8	92.3	86.4	95.8	94.8	98.5	96.2	96.1	96.6
Classification by chance (%) <sup>(a)</sup>	79.1	77.2	84.8	83.3	85.8	76.7	92.0	90.2	97.1	92.7	92.5	93.4

Notes: This table summarizes the qualitative findings of the study, where “-” means that the factor is not influent in the model; “↑” means that the factor impact is likely to increase the probability of occurrence of a work zone crash involving a certain contributing factor, ↑\* and ↑\*\* are the two largest impact factors that likely increase the probability; “↓” means that the factor impact is likely to decrease the probability of occurrence; ↓\* is the smaller impact factors that likely decrease the probability and “n.a.” means that the factor was not considered in the model. <sup>(a)</sup> Proportional chance criterion.

Table 8. Summary of model qualitative findings for driver age groups.

Driver Age Group	Under 25 Years Old			Between 25 and 64 Years Old			Over 64 Years Old		
Sample Size	358	267	91	1951	1372	579	288	197	91
Factor	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural
Urban	-	n.a.	n.a.	-	n.a.	n.a.	-	n.a.	n.a.
HV involved	-	-	-	↑*	↑*	-	↓*	-	-
Moto involved	-	-	-	-	-	-	-	-	-
Driver action	-	-	-	↑	↑	-	↑	-	↑**
Excessive speed	↑*	↑*	-	-	-	-	-	-	-
Obstacle	-	↓	-	↑**	↑**	-	-	-	-
Vertical signs	-	-	-	-	-	↓*	↑*	-	↑*
Safety distance	-	-	↑*	-	-	-	-	-	-
Speed limit	-	-	-	-	-	-	-	-	-
Intersection	-	-	-	-	↑	-	-	-	-
Horizontal design	-	-	↓*	-	-	-	-	-	-
Vertical design	-	-	-	-	-	-	-	-	-
Pav. Grip conditions	↓	↓	-	-	-	-	↑	-	↑
Luminosity	↓*	↓*	-	↑	↑	-	↑**	↑*	↑
Weather	-	-	-	-	-	-	-	-	-
Hosmer and Lemeshow Test	0.752	0.189	0.899	0.807	0.828	1.000	0.991	1.000	0.795
Stepwise Number of steps	12	9	12	12	9	13	11	12	10
Cases correctly classified (%)	86.6	85.9	88.2	73.1	72.4	75.4	89.5	89.6	88.5
Classification by chance (%) <sup>(a)</sup>	76.8	75.8	79.2	60.7	60.0	62.5	80.7	81.4	79.2

Notes: This table summarizes the qualitative findings of the study, where “-” means that the factor is not influent in the model; “↑” means that the factor impact is likely to increase the probability of occurrence of a work zone crash involving a certain driver age group, ↑\*and ↑\*\* are the two largest impact factors that likely increase the probability; “↓” means that the factor impact is likely to decrease the probability of occurrence; ↓\* is the smaller impact factors that likely decrease the probability and “n.a.” means that the factor was not considered in the model. <sup>(a)</sup> Proportional chance criterion.

As no information about vehicle-kilometers traveled through the work zones or work zone layout (geometry) was available, models for the rural and urban road environment were developed in addition to the models obtained for the total number of crashes and drivers involved, to indirectly consider the exposure to crash. Work zone crashes in an urban environment were the most common in the database considered (73.5% of crashes).

#### 4. Discussion

An overall interpretation of model results is presented to provide a more comprehensive understanding of the major causal factors involved in Portuguese work zone crashes.

##### (a) Crash type results analysis

As some types of crashes are more common in a given road environment, urban and rural models have provided more reliable information on risk factors involved in each type of crash. The analysis revealed, as expected, that work zone crashes involving pedestrians occur essentially in urban areas and involve mainly passenger car vehicles (less likely to involve motorized or heavy vehicles). The set of significant factors identified (see Table 6) points to the need to improve work zones conditions for pedestrians and workers on foot. With regard to angle crashes, they are common in urban areas, especially in or near road intersections. As the disregard of vertical signs is a risk factor that increases the probability of angle crash occurrence, the results point to the need to reinforce signaling and speed control near intersections. In rural areas, special care with heavy vehicles approach to intersections must be also considered. In simple run-of-road crashes, motorcycles have been identified as a risk factor in both urban and rural work zones (see Table 6). Measures to control speed (although not identified by the analysis as a risk factor) and awareness and training of motorcycle drivers can be pointed out as mitigating measures that can be implemented. Rear-end accidents occurred both in rural and urban areas. The main risk factors identified, excessive speed and disregard for the safety distance, point to the need for measures to control speed and safety distance between vehicles, especially in rural work zones. It is important to note that for some models, according to [39,40], the sample size used in the analysis was small so results must be read carefully (e.g., pedestrian and angle crashes in the rural environment). Still, the adjustment of models is acceptable and, although the percentage of cases correctly classified (predictive accuracy of models) is higher than the random classification (between 2.6% and 13.1%), it is not at least 25% higher, as recommended in the literature [38].

##### (b) Contributing factor result analysis

The results obtained for the unexpected obstacle models showed that the involvement of motorcycles and lack of grip and lighting conditions in work zones are the most significant risk factors (see Table 7). Only luminosity was found to be significant in rural work zone crashes, however, the sample size used to build the model is small [39,40]. Although results for luminosity factor must be carefully considered and grip condition was not identified for rural areas, in view of the 3 models results (total, urban, and rural), it is possible to point out the need to improve work zone light and grip conditions. With regard to excessive speed, better speed management measures must be considered, especially in work zones with winding horizontal road design and a rural environment. The main risk factors identified were driver action (running straight), speed limit, and horizontal design (Table 7). Crashes involving disregard for vertical signs occurred essentially in urban areas, with heavy vehicle involvement and at intersections (risk factors). Once again and in line with the results obtained for crash type, these results point to the need for a careful treatment of intersections, especially with regard to the signaling of these areas. Finally, disregard for safety distance models revealed as main risk factors the involvement of passenger car vehicles (less likely to involve motorized or heavy vehicles) running in horizontal straight work zone sections and good pavement grip conditions (see Table 7), pointing to the need of speed and distance management measures reinforcement. Most

of the rural environment models were developed from small samples, which limits the model's adjustment to data (goodness-of-fit), case classification, and interpretation of results. To overcome these aspects, the use of a longer period of analysis should be considered (more than 3 years). For the remaining models fit is acceptable (no significant differences between actual and predicted values) and the percentage of cases correctly classified is higher than the random classification (between 1.4% and 9.7%).

#### (c) Driver age groups results analysis

Portuguese official information on drivers involved in crashes does not allow to identify which driver caused the accident. This aspect greatly limits the analysis, since the data used to build the age group models includes all drivers involved in the crash. Despite this limitation, from the results obtained it is possible to verify that the age group under 25 years old presents as main risk factors excessive speed and poor lighting conditions (see Table 8), factors already identified in previous models. The main risk factors identified for the 25 to 64 years old age group are the involvement of heavy vehicles and obstacles in the pavement. For drivers over 64 years old, disregard for vertical signs was the most relevant risk factor, especially in rural areas (Table 8). Actions to raise awareness and respect for speed limits, as well as adequate work zone lighting are measures to be considered to mitigate the identified young driver's risk factors. Convenient delimitation and signaling of work zones and work equipment are measures to be considered to overcome the risk factors identified for 25 to 64 and over 64 years old age groups. Regarding the stepwise process, the number of steps to reach the final models was in general higher than the ones obtained in the type of crash and contributing factor models. However, model goodness-of-fit test results were good and the cases correctly classified are superior to the random classification by 8.2 to 12.9%.

## 5. Conclusions

The current study aimed to analyze the data registered in Portuguese police crash reports, which include part of the information presented in Table 1, to assess the possibility of detecting risk factors involved in Portuguese road work zones crashes.

Since the occurrence of a crash is a binary response and it can be described by a set of explanatory variables, binary logistic and probit regressions were considered suitable statistical methods for the study of the available work zone crash data. The application of these methods requires special attention to the model building since some decisions will depend entirely on the expertise of the road safety researcher and may influence the final results. These decisions were carefully taken over the work zone model construction (model selection method, significance level, and goodness-of-fit test).

Despite the data limitations identified, binary logistic and probit regression analysis applied to the 2013–2015 Portuguese official crash data allowed to predict the occurrence of work zone crashes for most of the situations considered and to identify significant risk factors by type of crash, main contributing factor, and driver age group. The limitations identified, such as the number of reports with no clear information on the occurrence of the crash in a work zone (22.1% of crashes), other report fields filling failures, and the lack of information related to the number, length, duration, layout, operation, and type of work or traffic flow, points to the need for a more complete and clear information record to allow a more representative and reliable perception of all the risk factors involved and the evolution to more sophisticated statistical approaches.

To overcome some of these limitations, the authors decided to perform the analysis for the total number of work zone crashes and by road environment (urban and rural). Analysis by road environment allowed to indirectly consider some of the aspects not registered in the police reports, such as the work zone layout, traffic characteristics, and the most common roadworks, normally specific of each road environment.

Concerning the comparison between logit and probit link functions, the estimated coefficients by both approaches were very similar. Only the vertical signs and safety distance models presented a slight difference which may be related to the small sample

size available to develop the models (<100 cases). It is anticipated that the use of a long period in future analyzes (more than 3 years) and an adequate and complete crash report filling will result in larger samples and estimation of very similar coefficients for the two approaches.

Most of the risk factors identified in the 2013–2015 Portuguese work zone crash data modeling (by type of crash, contributing factor, and age group), for the urban and rural environment, are in line with previous research findings [6,8,10–13].

In summary, results support the need to improve:

- (a) Work zone signaling and traffic speed control in the approach and throughout work zones, in Portuguese rural roads.
- (b) Work zone signaling, lighting, and delimitation; traffic speed control; special treatment of road intersections; and specific measures targeted at motorcycles and heavy vehicle drivers, inside Portuguese urban areas.

Finally, while police crash reports are not revised to incorporate more detailed work zone crash information and until the promotion of proper filling of reports begins to take effect, changes in risk factors that are currently possible to identify and their influence on the occurrence of work zone crashes, can be detected by analyzing time series of crash records over time. The identification of these changes can be used to assess the efficacy of measures adopted to minimize the occurrence of work zone crashes and to support the definition of new measures, analysis approaches, and work zone layout design.

**Author Contributions:** Conceptualization, B.S. and L.P.-S.; methodology, B.S., V.T., C.P. and L.P.-S.; validation, B.S. and V.T.; formal analysis, B.S. and V.T.; writing—original draft preparation, B.S. and V.T.; writing—review and editing, B.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of data. Data was obtained from Portuguese Road Safety Authority (ANSR—Autoridade Nacional de Segurança Rodoviária) and are available from the corresponding author if authorized by ANSR.

**Acknowledgments:** The authors acknowledge the Portuguese Road Safety Authority (ANSR—Autoridade Nacional de Segurança Rodoviária) for providing data and assistance, CERIS—Civil Engineering Research and Innovation for Sustainability (ECI/04625), and the University of Beira Interior for support and funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ullman, G.L.; Ullman, B.R.; Finley, M.D. Analysis of Crashes at Active Night Work Zones. In Proceedings of the TRB 85th Annual Meeting, Washington, DC, USA, 22–26 January 2006. Transportation Research Board.
2. Meng, Q.; Weng, J.; Qu, X.-B. A probabilistic quantitative risk assessment model for the long-term work zone crashes. *Accid. Anal. Prev.* **2010**, *42*, 1866–1877. [[CrossRef](#)] [[PubMed](#)]
3. Silverstein, C.; Schorr, J.; Hamdar, S.H. Work zones versus nonwork zones: Risk factors leading to rear-end and sideswipe collisions. *J. Transp. Saf. Secur.* **2015**, *8*, 310–326. [[CrossRef](#)]
4. La Torre, F.; Domenichini, L.; Nocentini, A. Effects of stationary work zones on motorway crashes. *Saf. Sci.* **2017**, *92*, 148–159. [[CrossRef](#)]
5. Ambros, J.; Turek, R.; Elgner, J.; Křivánková, Z.; Valentová, V. Effectiveness Evaluation of Section Speed Control in Czech Motorway Work Zones. *Safety* **2020**, *6*, 38. [[CrossRef](#)]
6. Pigman, J.G.; Agent, K.R. *Analysis of Accidents in Construction and Maintenance Work Zones*; Research Report UKTRP-88-13; Kentucky Transportation Cabinet: Lexington, KY, USA, 1988.
7. Garber, N.J.; Zhao, M. Distribution and Characteristics of Crashes at Different Work Zone Locations in Virginia. *Transp. Res. Rec. J. Transp. Res. Board* **2002**, *1794*, 19–25. [[CrossRef](#)]

8. Bai, Y.; Li, Y. *Determining Major Causes of Highway Work Zone Accidents in Kansas*; Final Report; Kansas Department of Transportation/Bureau of Materials and Research: Topeka, KS, USA, 2006.
9. Swedish Transport Administration (Trafikverket). Plötsligt Var Det Ett Vagarbete! En Studie Av Trafikolyckor Vid Vagarbeten 2003–2009 Med Speciellt Fokus På Upphinnandeolyckor. Available online: [https://trafikverket.ineko.se/Files/sv-SE/11501/RelatedFiles/2011\\_007\\_plotsligt\\_var\\_det\\_ett\\_vagarbete.pdf](https://trafikverket.ineko.se/Files/sv-SE/11501/RelatedFiles/2011_007_plotsligt_var_det_ett_vagarbete.pdf) (accessed on 2 October 2020).
10. Santos, B.; Picado-Santos, L. Impact Assessment of Road Work Zones in Operating Conditions, Safety and User Costs. Pavement and Asset Management. In Proceedings of the World Conference on Pavement and Asset Management, Baveno, Italy, 12–16 June 2017; Maurizio, C., Ed.; CRC Press/Balkema Taylor & Francis Group: Boca Raton, FL, USA, 2017.
11. Li, Y.; Bai, Y. Effectiveness of temporary traffic control measures in highway work zones. *Saf. Sci.* **2009**, *47*, 453–458. [[CrossRef](#)]
12. Paolo, P.; Sar, D. Driving Speed Behaviour Approaching Road Work Zones On Two-Lane Rural Roads. *Procedia Soc. Behav. Sci.* **2012**, *53*, 672–681. [[CrossRef](#)]
13. Osman, M.; Paleti, R.; Mishra, S.; Golias, M. Analysis of injury severity of large truck crashes in work zones. *Accid. Anal. Prev.* **2016**, *97*, 261–273. [[CrossRef](#)] [[PubMed](#)]
14. FORMAT—Fully Optimised Road Maintenance. *Project, Final Technical Report*; Project funded by the European Community; Dienst Weg- en Waterbouwkunde, Ministerie van Verkeer en Waterstaat, Directoraat-Generaal Rijkswaterstaat: Delft, The Netherlands, 2005.
15. ARROWS—Advanced Research on Road Work Zone Safety Standards in Europe. *Project, Deliverable 4, Volume II, Road Work Zone Safety Practical Handbook*; Project funded by the European Community; National Technical University of Athens: Athens, Greece, 1998.
16. *Motor Vehicle Traffic Accident Police Investigation Report*; MV6020; BC, Canada, 1998.
17. Investigating Agency. *Kentucky Uniform Police Traffic Collision Report*; KSP 74 (1/00); Investigating Agency: Frankfort, KY, USA, 2012.
18. Massachusetts Department of Transportation. *Motor Vehicle Crash Operator Report*; 2012–2020 Report; CRA-23; Massachusetts Department of Transportation: Boston, MA, USA, 2012.
19. Texas Department of Public Safety. *Texas Peace Officer's Crash Report*; CR-3CS; Texas Department of Public Safety: Austin, TX, USA, 2010.
20. Autoridade Nacional de Segurança Rodoviária (ANSR). *Boletim Estatístico de Acidentes de Viação*; Ministério da Administração Interna: Lisbon, Portugal, 2013.
21. Dirección General de Tráfico. *Formulario de Accidentes con Víctimas*; Boletín Oficial del Estado, n. 289; Ministerio del Interior: Montevideo, Spain, 2014.
22. Istituto Nazionale di Statistica (ISTAT). *Rilevazione degli Incidenti Stradali con Lesioni a Persone*; Istituto Nazionale di Statistica (ISTAT): Rome, Italy, 2016.
23. *STATS19 Road Accident Injury Statistics*; MG NSRF/A; Reporting Form: UK, 2011.
24. Pigman, J.G.; Agent, K.R. Highway Accidents in Construction and Maintenance Work Zones. *Transp. Res. Rec. J. Transp. Res. Board* **1990**, *1270*, 12–21.
25. Dissanayake, S.; Akepati, S. *Identification of Work Zone Crash Characteristics. Research Report*; Sponsored by the Smart Work Zone Deployment Initiative and the Iowa Department of Transportation; Iowa State University: Ames, IA, USA, 2009.
26. Charman, S.; Fletcher, J.; Lawton, B.; Scoons, J.; Clark, S.; Cocu, X.; Nuallain, N.; Turk, M.; Zimmermann, M.; Rillie, I. BRoWSER: Base-Lining Road Works Safety on European Roads. Final trial report, Project funded under the CEDR Transnational Road. 2014. Available online: <https://www.cedr.eu/strategic-plan-tasks/research/cedr-call-2012/call-2012-safety/browser-project-results/> (accessed on 2 October 2020).
27. Yang, H.; Ozbay, K.; Ozturk, O.; Xie, K. Work Zone Safety Analysis and Modeling: A State-of-the-Art Review. *Traffic Inj. Prev.* **2014**, *16*, 387–396. [[CrossRef](#)] [[PubMed](#)]
28. Osman, M.; Paleti, R.; Mishra, S. Analysis of passenger-car crash injury severity in different work zone configurations. *Accid. Anal. Prev.* **2018**, *111*, 161–172. [[CrossRef](#)] [[PubMed](#)]
29. Lu, J.; Wang, Z.; Wang, X. *Integrated Work Zone Safety Management System and Analysis Tools*; Report; Florida Department of Transportation: Tallahassee, FL, USA, 2008.
30. Mannering, F.L.; Bhat, C.R. Analytic methods in accident research: Methodological frontier and future directions. *Anal. Methods Accid. Res.* **2014**, *1*, 1–22. [[CrossRef](#)]
31. Santos, B.; Picado-Santos, L.; Trindade, V. Using Binary Logistic Regression to Explain the Impact of Accident Factors on Work Zone Crashes. In Proceedings of the RSS 2017—Road Safety & Simulation, Delft, The Netherlands, 17–19 October 2017.
32. Portuguese National Authority of Road Safety (ANSR). Available online: <http://www.ansr.pt> (accessed on 2 October 2020).
33. Trindade, V. Road Work Zone Crash Data Analysis (in Portuguese). Master's Thesis, University of Beira Interior, Covilhã, Portugal, 2017.
34. Polonia, C. Registration and Diagnosis of Road Accidents in Work Zones (in Portuguese). Master's Thesis, University of Beira Interior, Covilhã, Portugal, 2014.
35. SAS Institute Inc. *SAS/STAT®13.2 User's Guide*; SAS Institute Inc.: Cary, NC, USA, 2014.
36. IBM Corp. *IBM SPSS Statistics for Windows*, version 24.0; IBM Corp.: Armonk, NY, USA, 2016.

37. O'Halloran, S. *Lecture 9: Logit / Probit. Sustainable Development U9611 Econometrics II*; Columbia University: New York, NY, USA, 2016.
38. Marôco, J. *Análise Estatística com o SPSS Statistics*, 7th ed.; ReportNumber: Pêro Pinheiro, Portugal, 2018; ISBN 9789899676350.
39. Long, J. *Regression Models for Categorical and Limited Dependent Variables*; SAGE Publications: Thousand Oaks, CA, USA, 1997.
40. Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T.R.; Feinstein, A.R. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **1996**, *49*, 1373–1379. [[CrossRef](#)]
41. Freitas, L.; Filho, S.; Júnior, J.; Silva, F. Comparison of Logit and Probit Link Functions in Binary Regression Considering Different Sample Sizes. *Enciclopédia Biosfera.* **2013**, *9*, 2936–2951.
42. Hosmer, D.; Lemeshow, S.; Sturdivant, R. *Applied Logistic Regression*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013.